Une preuve "moderne" du théorème de D´Alembert-Gauss

Dominique Hoareau, domeh@wanadoo.fr

Position du problème :

P désigne un polynôme non constant de $\mathbb{C}[X]$, de degré $n \geqslant 1$. On veut prouver que P admet une racine complexe. La preuve proposée met en jeu une propriété topologique simple de $\mathbb{C}$: la connexité de $\mathbb{C}$ privé d´un ensemble fini de points (résultat grossièrement faux dans $\mathbb{R}$).

Pour tout $z = x + iy \in \mathbb{C}$, $P(z)$ s´écrit : $P(z) = P(x + iy) = Q(x, y) + iR(x, y)$ où Q et R sont des fonctions polynômes en x et y, à valeurs dans $\mathbb{R}$. On note f la fonction biréelle associée à P, définie sur $\mathbb{R}^2$ par $f(x, y) = (Q(x, y), R(x, y))$. L´application f est clairement C^1 comme ses composantes polynômiales.

1- Pour prendre un bon départ

Résultat 1 : Si P s´écrit : $P(X) = a_0 + a_1X + a_2X^2 + ... + a_nX^n$, on appelle polynôme dérivé de P le polynôme $P'(X) = a_1 + 2a_2X + ... + na_nX^{n-1}$. On a :

$$\forall (x, y) \in \mathbb{R}^2 \quad \frac{\partial Q}{\partial x}(x, y) = \frac{\partial R}{\partial y}(x, y) = \Re\big(P'(x + iy)\big) \quad ; \quad -\frac{\partial Q}{\partial y}(x, y) = \frac{\partial R}{\partial x}(x, y) = \Im\big(P'(x + iy)\big).$$

Preuve : Pour $z \in \mathbb{C}$, la famille $\big(1, X - z, (X - z)^2, ..., (X - z)^n\big)$ est une base de $\mathbb{C}^n[X]$, donc $P(X)$ s´écrit : $P(X) = \alpha_0 + ... + \alpha_n(X - z)^n$. Pour déterminer $\alpha_k \quad (0 \leqslant k \leqslant n)$, on dérive k fois les deux membres de l´égalité et on fait $X = z$. On obtient la formule (exacte) de Taylor :

$$P(X) = P(z) + P'(z)(X - z) + ... + \frac{P^{(n)}(z)}{n!}(X - z)^n.$$

On écrit alors pour $(h, k) \in \mathbb{R}^2$: $P(x+iy+h+ik) = P(x+iy)+P'(x+iy)(h+ik)+ \mid h+ik \mid \varepsilon(h+ik)$ avec $\lim_{h+ik\to 0}\varepsilon(h + ik) = 0$. Ce "développement limité" de P au premier ordre en tout $x + iy$ donne après séparations en parties réelle et imaginaire :

- $Q(x + h, y + k) = Q(x, y) + h\Re(P'(x + iy)) - k\Im(P'(x + iy))+ \mid h + ik \mid \Re(\varepsilon(h + ik))$
- $R(x + h, y + k) = R(x, y) + h\Im(P'(x + iy)) + k\Re(P'(x + iy))+ \mid h + ik \mid \Im(\varepsilon(h + ik))$.

Par unicité des différentielles de Q et R en (x, y), il vient :

$$\forall (x, y) \in \mathbb{R}^2 \quad \frac{\partial Q}{\partial x}(x, y) = \frac{\partial R}{\partial y}(x, y) = \Re\big(P'(x + iy)\big) \quad ; \quad -\frac{\partial Q}{\partial y}(x, y) = \frac{\partial R}{\partial x}(x, y) = \Im\big(P'(x + iy)\big).$$

Grâce à ces formules dites de Cauchy-Riemann, on prouve que le jacobien de f en (x, y) (déterminant de la matrice de $df_{(x,y)}$ dans la base canonique de $\mathbb{R}^2$) vaut $\mid P'(x + iy) \mid^2$. Ainsi, $df_{(x,y)}$ est inversible si et seulement si $x + iy$ n´est pas un zéro de P'.

Résultat 2 : Si D désigne une partie dénombrable de $\mathbb{R}^2$ (ou de $\mathbb{C}$), alors $\mathbb{R}^2 \setminus D$ (ou $\mathbb{C} \setminus D$) est connexe par arcs.

Preuve : Soit : A, B deux points de $\mathbb{R}^2 \setminus D$. On choisit un repère de $\mathbb{R}^2$ d´origine A et tel que $B(1,0)$. Pour $k \in \mathbb{R}$ et $x \in [0,1]$, on pose : $f_k(x) = kx(1-x)$. On note G_k le graphe de f_k. Il existe $k_0 \in \mathbb{R}$ tel que $G_{k_0} \cap D = \oslash$. Sinon, $\forall k \in \mathbb{R} \quad \exists M_k \in G_k \cap D$ avec nécessairement $M_k \neq A$ et $M_k \neq B$. On envisage alors l´application $\iota : k \in \mathbb{R} \mapsto M_k \in D$. Si k_1 et k_2 sont deux réels tels que $\iota(k_1) = (x_1, y_1)$ et $\iota(k_2) = (x_2, y_2)$ sont confondus, alors $y_1 = y_2$ donne $k_1 x_1(1-x_1) = k_2 x_2(1-x_2)$ et puisque $x_1 \neq 0$, $1 - x_1 \neq 0$ et $x_1 = x_2$, il vient $k_1 = k_2$. L´application ι est donc injective, $\mathbb{R}$ est donc dénombrable. Contradiction.

Résultat 3 : Un polynôme de degré $d \geqslant 1$ à coefficients dans un corps $\mathbb{K}$ a au plus d racines.

Remarque : Ce résultat est faux lorsqu´on remplace le corps $\mathbb{K}$ par un anneau : considérer le polynôme $4X$ de degré 1, à coefficient dans $\mathbb{Z}/8\mathbb{Z}$, qui a 3 racines 0,2 et 4.
Conséquence : un polynôme de $\mathbb{C}[X]$ qui a une infinité de zéros, est nécessairement constant.

2- Un outil : inversion locale

(cf Cours d´analyse, A. Pommelet, Ellipses)

Soit I un intervalle de $\mathbb{R}$, a un point intérieur de I et $f : I \to \mathbb{R}$ une application de classe C^1 sur I. On suppose que $f'(a) \neq 0$ ($f'(a) > 0$ pour fixer les idées). Par continuité de f' en a, on choisit $r > 0$ tel que :

$$]a-r, a+r[\subset I \quad ; \quad \forall x \in]a-r, a+r[\; f'(x) > 0.$$

Ainsi f est strictement croissante sur $]a-r, a+r[$ (conséquence du théorème des accroissements finis TAF), donc f est injective sur $]a-r, a+r[$ (ordre total sur $\mathbb{R}$ et conservation de l´ordre par f), donc f réalise une bijection de $]a-r, a+r[$ sur son image.

On généralise ce résultat à $\mathbb{R}^p$ pour des fonctions de plusieurs variables. La condition "$f'(a) \neq 0$" est remplacée par "df_a inversible". L´ordre sur $\mathbb{R}$ et le TAF ne sont plus d´actualité, les nouveaux arguments sont "complétude de $\mathbb{R}^p$" et inégalité des accroissements finis vectoriels IAF.

Enoncé : Soit U un ouvert de $\mathbb{R}^p$, $a \in U$ et $f : U \to \mathbb{R}^p$ de classe C^1. On suppose que df_a est un automorphisme linéaire de $\mathbb{R}^p$. Alors, il existe un voisinage ouvert B de a et un voisinage ouvert $\widetilde{B}$ de $f(a)$ tel que f réalise une bijection de B sur $\widetilde{B}$.
Ce résultat est intuitif (mais loin d´être trivial) : sur un voisinage de a, f et sa meilleure approximation affine $x \mapsto f(a) + df_a.(x-a)$ autour de a ont des comportements proches ; df_a étant inversible, f est aussi bijective au voisinage de a.

Lemme topologique : Soit $a \in \mathbb{R}^p$, $r > 0$ et $g : B(a,r) \to \mathbb{R}^p$ une k-contraction ($0 < k < 1$) qui envoie a sur 0. Si on perturbe l´identité de $B(a,r)$ par g en considérant $f : x \mapsto x + g(x)$, alors $B(a,r)$ est réduite par f de façon bijective en la boule $B\left(a, (1-k)r\right)$.

L´injectivité de f est claire : si $f(x)=f(y)$, alors $g(y)-g(x)=x-y$ et on conclut aisément que $x=y$ puisque g est k-lipschitzienne avec $0<k<1$.
Soit à présent $y\in B\left(a,(1-k)r\right)$. Pour $x\in B(a,r)$, $y=f(x)\Leftrightarrow h(x)=x$ où $h(x)=y-g(x)$. On se ramène donc à la recherche d´un point fixe pour h dans $B(a,r)$ (nouveau problème plus naturel quand on remarque que h est comme g, k-contractante). Un schéma d´approximations successives (à la Picard) semble s´imposer. La preuve demande du soin puisque h ne stabilise pas à priori la boule $B(a,r)$. Par ailleurs, le point fixe qui sera obtenu par itérations, appartient à priori à l´adhérence de $B(a,r)$ et non pas à $B(a,r)$.
On pose : $x_0=a$. On a : $\| h(x_0)-a \|=\| h(x_0)-x_0 \|=\| y-a \|<(1-k)r$ donc $h(x_0)\in B(a,r)$.
On pose : $x_1=h(x_0)$. Encore : $\| h^2(x_0)-h(x_0) \|\leqslant k \| h(x_0)-x_0 \|<k(1-k)r$ d´où par inégalité triangulaire : $\| h^2(x_0)-a \|<(1+k)\,(1-k)r=(1-k^2)r$, ce qui assure que : $x_2=h^2(x_0)\in B(a,r)$.
On suppose construits $x_0,x_1,x_2,...,x_n$ dans $B(a,r)$ avec pour $1\leqslant i\leqslant n$: $\| x_i-x_{i-1} \|<k^{i-1}(1-k)r$.
On a : $\| h(x_n)-x_n \|\leqslant k \| x_n-x_{n-1} \|<k\, k^{n-1}(1-k)r=k^n(1-k)r$ donc par inégalités triangulaires : $\| h(x_n)-x_0 \|<(1-k)r\ \frac{1-k^{n+1}}{1-k}=(1-k^{n+1})r$. On définit donc correctement une suite de $B(a,r)$ en partant de $x_0=a$ et de la relation $x_{n+1}=h(x_n)$. L´inégalité $\| x_{n+1}-x_n \|<k^n(1-k)r$ assure que la série numérique $\sum \| x_{n+1}-x_n \|$ converge et $\sum_{n=0}^{\infty} \| x_{n+1}-x_n \|<\frac{(1-k)r}{1-k}=r$. Puisque $\mathbb{R}^p$ est complet, la série télescopique $\sum x_{n+1}-x_n$ et donc (x_n) convergent (vers un vecteur noté x) dans l´adhérence de $B(a,r)$. Mais puisque $\| x \|\leqslant \sum_{n=0}^{\infty} \| x_{n+1}-x_n \|<r$, en fait : $x\in B(a,r)$. Enfin, par continuité de h, $x=h(x)$.

Fin de la démonstration : Par pure commodité, on réduit la preuve en remplaçant f par :

$$\underbrace{df_a^{-1}\circ f}_{on\,peut\,supposer\,désormais\,:\,df_a=Id} + \underbrace{a-df_a^{-1}.f(a)}_{et\,:\,f(a)=a}.$$

L´idée est d´écrire sur une boule ouverte centrée en a : $f=Id+g$ où g est une contraction qui envoie a sur 0, puis on évoque le lemme topologique.
Avec la continuité de $z\mapsto df_z$, on choisit $r>0$ tel que :

$$\forall z\in B(a,r)\subset U \quad ||| \, df_z-Id \, |||\leqslant \frac{1}{2} \quad ✠.$$

Pour $x,y\in B(a,r)$, on souhaite évaluer : $\| (f-Id)(y)-(f-Id)(x) \|$. Pour $t\in[0,1]$, on pose : $\alpha(t)=f(x+t(y-x))-[x+t(y-x)]$. L´application α est dérivable sur $[0,1]$ avec

$$\alpha'(t)=df_{x+t(y-x)}.(y-x)-(y-x)$$

donc $\| \alpha'(t) \|\leqslant ||| \, df_{x+t(y-x)}-Id \, ||| \ \| y-x \|\leqslant \frac{1}{2} \| y-x \|$ par convexité de $B(a,r)$ et ✠. Avec l´IAF,

$$\| \alpha(1)-\alpha(0) \|=\| (f(y)-y)-(f(x)-x) \|\leqslant \frac{1}{2} \| y-x \|.$$

Ainsi, $g=f-Id$ est $\frac{1}{2}$-contractante sur $B(a,r)$ et envoie a sur 0, donc avec le lemme topologique, $f=Id+g$ réalise une bijection de $B(a,r)$ sur $B(a,\frac{1}{2}r)$.

3- La preuve de Milnor (1965)

(cf Exercice 82, Les cahiers de la RMS6, Vuibert
et Introduction aux variétés différentielles, J. Lafontaine, EDP Sciences, page 58)

Notations : Soit $\mathcal{C}_f$ l´ensemble des zéros de P', en nombre au plus égal à $d^\circ(P') = n-1$. $\mathcal{C}_f$ est appelé ensemble des points critiques de f. On pose :

- $\mathcal{V}_f = f(\mathcal{C}_f)$ (Ensemble des valeurs critiques de f)
- $V = \mathbb{R}^2 \setminus \mathcal{V}_f$
- $U = f^{-1}(V) = \mathbb{R}^2 \setminus \mathcal{C}_f$.

V est un ouvert connexe (par arcs) de $\mathbb{R}^2$ puisque la partie $\mathcal{V}_f$ est finie et U est ouvert par continuité de f. Si $(0,0) \in \mathcal{V}_f$, $(0,0)$ est une valeur prise par f, donc 0 une valeur prise par P comme souhaité. On suppose désormais que $(0,0) \in V$.

On introduit la fonction ν, à valeurs dans $\mathbb{N}$ définie sur V par $\nu((X,Y)) = \text{card}\{f^{-1}\{(X,Y)\}\} \leqslant n$. Soit $(X_0,Y_0) \in V$ et $p = \nu(X_0,Y_0)$.

Objectif : Montrer que ν est constante sur V de valeur p.

Sratégie : Montrer que ν est localement constante, donc continue. La source V étant connexe et le but $\mathbb{N}$ discret, ν est alors constante.

a) On suppose que $p = \nu(X_0,Y_0) \geqslant 1$. Les éléments de $f^{-1}\{(X_0,Y_0)\}$ étant les points $z_k = a_k + ib_k \quad (1 \leqslant k \leqslant p)$ de l´ouvert U, par inversion locale, il existe des voisinages ouverts $\widetilde{v_k}$ de (X_0,Y_0), des voisinages ouverts v_k de z_k tels que f réalise localement une bijection de v_k sur $\widetilde{v_k}$. Dans $\mathbb{R}^2$ séparé, quitte à restreindre les v_k, on peut les supposer disjoints deux à deux. On considère le voisinage ouvert $v = \bigcap_{1\leqslant k\leqslant p} \widetilde{v_k}$ de (X_0,Y_0). Quitte à changer v_k par l´ouvert $v_k \cap f^{-1}(v)$, f réalise localement une bijection de chaque v_k sur v. Clairement, pour tout $(X,Y) \in v$, $\nu(X,Y) \geqslant p$.

b) Quitte à restreindre v, on montre que : $\forall (X,Y) \in v \ , \quad \nu(X,Y) \leqslant p$. Par l´absurde, pour tout $m \in \mathbb{N}^*$, il existe (X_m,Y_m) dans $v \cap B((X_0,Y_0),\frac{1}{m})$ admettant un antécédent (x_m,y_m) qui ne soit dans aucun v_k. Puisque $\lim_{|z|\to+\infty} \mid P(z) \mid = +\infty$, $\exists R > 0 \quad \mid z \mid > R \Rightarrow \mid P(z) \mid \geqslant \| (X_0,Y_0) \| +1$, ce qui permet de dire que la suite (x_m,y_m) est à valeurs dans le fermé borné $\overline{B(0,R)}$ de $\mathbb{R}^2$. Sinon, pour un certain $m \in \mathbb{N}^*$, $\| (x_m,y_m) \| > R$, $\| (X_m,Y_m) \| = \mid P(x_m+iy_m) \mid \geqslant \| (X_0,Y_0) \| +1$, ce qui n´est pas puisque $(X_m,Y_m) \in B((X_0,Y_0),1)$. A présent, on extrait à bon droit de la suite (x_m,y_m), une sous-suite qui converge (dans le fermé $\mathbb{R}^2 \setminus \bigsqcup_{1\leqslant k\leqslant p} v_k$) vers un (x,y). Par continuité de f, on a $f(x,y) = (X_0,Y_0)$ avec $(x,y) \notin \bigsqcup_{1\leqslant k\leqslant p} v_k$, donc $\nu(X_0,Y_0) > p$. Absurde.

c) Si $\nu(X_0,Y_0) = 0$, comme en b), on montre que ν est nulle sur un voisinage de (X_0,Y_0).

ν est donc localement constante autour de (X_0,Y_0), (X_0,Y_0) étant choisi arbitrairement dans V, donc ν est constante sur V par un argument de connexité. Or on peut toujours choisir (X_0,Y_0) de sorte que $p \neq 0$. Sinon, $\forall (x,y) \in \mathbb{R}^2 \quad f((x,y)) \in \mathcal{V}_f$, une valeur critique c est alors atteinte une infinité de fois par P, ce qui donne une infinité de zéros pour le polynôme $P-c$, donc P est constant. Contradiction.

Bilan : Tout élément de V est atteint p fois par f (avec $p > 0$), en particulier, 0 est une valeur prise par P et on a montré que le corps $\mathbb{C}$ est algébriquement clos.